

DOCUMENT RESUME

ED 246 073

TM 840 344

AUTHOR Melican, Gerald J.; Flake, Barbara S.
TITLE Are Correction for Guessing and Nedelsky's Standard Setting Method Compatible?
PUB DATE Apr 84
NOTE 19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 24-26, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cutting Scores; Guessing (Tests); Higher Education; Multiple Choice Tests; Quantitative Tests; *Response Style (Tests); *Scoring Formulas; Testing Problems; *Test Wiseness; Undergraduate Students
IDENTIFIERS *Correction for Guessing; *Nedelsky Method

ABSTRACT

The validity of combining a correction for guessing with the Nedelsky-based cutscore was investigated. A five option multiple choice Mathematics Achievement Test was used in the study. Items were selected to meet several criteria. These included: the capability of measuring mathematics concepts related to performance in introductory statistics; the requirement of a maximum of high school level mathematics to complete the question; and items which made up a test with satisfactory reliability. Results indicated that minimally competent examinees tended to omit those items identified as "omit" items based on judges' ratings. The study suggested that the effect of applying the correction for guessing to a Nedelsky cutscore may not be too severe. Instructions which require examinees to physically eliminate alternatives that they know are incorrect for each item may decrease the tendency for some examinees to omit items for which they have partial knowledge. This change in test taking behavior of some examinees may increase the range of minimally competent candidates for whom the corrected for guessing adjusted cutscore is appropriate. (DWH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED246073

Method

1

Are Correction for Guessing and Nedelsky's
Standard Setting Method Compatible?

Gerald J. Melican

Educational Testing Service

Barbara S. Plake

University of Nebraska-Lincoln

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

Paper Presented at the Annual Meeting of the
National Council on Measurement in Education
New Orleans, 1984

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

G. J. Melican

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Running Head: STANDARD SETTING METHOD

Are Correction for Guessing and Nedelsky's
Standard Setting Method Compatible?

When determining the cutscore for a multiple choice examination using a judgment-based method, a panel of subject matter experts usually reviews each item in the examination. With the Nedelsky (1954) method, for example, subject matter experts independently indicate which of the incorrect responses of an item the "minimally competent" candidate will be able to recognize as incorrect. It is assumed that the examinee will guess from the remaining options. That is, if an examinee is able to eliminate one option as incorrect on a five option item, then the examinee has a probability of .25 of answering the item correctly. Each judge's individually determined cutoff score is then established by summing the probabilities over all the items. This represents the minimally competent test taker's expected score based on this judge's rating of these items. These expected scores for the minimally competent candidate are then averaged for the set of judges.

Occasionally, the examination used in assessing examinee competence or achievement is scored using a correction for guessing formula. When this is the case, Livingston and Zieky (1982) recommend applying a correction for guessing adjustment to the individual judges' previously determined cutscore prior to averaging. This correction for guessing adjustment reduces the

minimally competent test taker's expected score by subtracting from it a fraction of the expected number of wrong answers.

Combining a correction for guessing formula with a cutscore determined by a judgment-based standard setting method potentially introduces a problem due to the potential insensitivity of these techniques to examinee test taking behaviors. This is essentially only true with the Nedelsky method because the judges focus directly on the alternatives that a minimally competent candidate will be able to eliminate, and the probability of the test taker getting the item correct is established as a function of these ratings. Thus, the Nedelsky cutscore method prohibits the judges including examinee test taking behavior in their individually determined cutscores. In other judgment-based systems, (e.g., Angoff, 1971), the judges directly provide the probability that the minimally competent candidate will get the item correct and this subjective assessment may incorporate other test taking behaviors of the test taker other than the elimination of options for which the examinee is expected to have partial knowledge. The Livingston and Zieky correction when applied to the Nedelsky cutscore method assumes that a minimally competent candidate omits only those items for which the examinee could not eliminate any of the options. If, however, examinees omit items for which they have partial knowledge (i.e., items for which they can identify one or more of the item's options as incorrect), the application

of the correction for guessing formula to the Nedelsky cutscore will result in a corrected cutscore that is too stringent. The algebraic proof of this point is shown in Figure 1. The correction for guessing adjustment and the Nedelsky standard setting method, then, are only compatible if minimally competent candidates respond to all questions for which they have partial knowledge (i.e., only omit items for which they cannot eliminate any of the options).

The purpose of this study was to examine whether minimally competent candidates do in fact only omit those items for which the judges indicate these examinees would be unable to eliminate any of the options. Through this analysis, the validity of combining a correction for guessing to the Nedelsky-based cutscore was investigated.

Method

Instrument

The study used a 48-item five-option multiple-choice examination, Mathematics Achievement Test (MAT), which measured the mathematics skills and competencies prerequisite to satisfactory performance in an entry graduate level Introduction to Statistics class. The items were selected from a previously developed item pool to meet the following criteria:

1. They had to measure mathematics concepts related to performance in introductory statistics (Baggaley, 1969).

2. A maximum of high school level of mathematics was required to complete the question.

3. The items comprised a test with satisfactory reliability. Item difficulty indices for the items selected for the MAT ranged from 0.15 to 0.84 with a median value of 0.55; item discrimination indices (point biserial) ranged from 0.15 to 0.82 with a median of 0.53. The KR(20) reliability of the instrument was calculated to be 0.91.

The examination scores were used to identify student who would benefit from a remedial mathematics lab.

Panel of Judges

The expert judges were five professors who regularly teach a section of Introduction to Statistics in a College of Education at a large midwestern university. They all were familiar with the typical student who enrolls in one of the multiple sections of Introduction to Statistics and with the level of mathematics knowledge required of a minimally competent student in the course. These judges independently rated each item of the 48-item Mathematics Achievement Test, using the Nedelsky procedure to identify and circle the distractors (incorrect options) that a minimally competent candidate would be able to eliminate as the possible keyed answer.

Subjects and Administration

The Mathematics Achievement Test was administered to all students in Introduction to Statistics classes by their instructors during the first week of the Fall, 1983 semester. A total of 108 students took the examination in the four sections of the course. Written instructions to the examinees included the following:

Your score on this examination will take into account your incorrect as well as your correct answers. That is, the number of items you answer correctly will be reduced by subtracting one quarter of the number of items answered incorrectly. Therefore, to maximize your score, if you cannot eliminate any alternatives for a particular item, you should not guess the answer but rather leave that item blank. However, if you can eliminate one or more alternatives as incorrect for a particular item, it is to your advantage to select an answer from the remaining set.

These directions were read aloud to the students by their instructors. Further, the students were informed that if they scored below a cutoff score they would be assigned to participate in the mathematics remediation lab.

Procedure

Items on the examination for which two or more judges indicated that minimally competent examinees would be able to eliminate none

of the options were deemed "omit items" and the remaining items were labeled "non-omit" items.

Those examinees whose corrected for guessing examination scores fell within two standard errors of measurement of the adjusted cutoff score were considered the "minimally competent candidates." Their omit pattern on the 48-item instrument was analyzed. The percent of minimally competent examinees who omitted each item was calculated and these percentages were ranked. The Mann-Whitney U test was performed on the ranks with the judges ratings of the items (omit/non-omit) as the grouping factor.

In addition, items that 50% or more of the examinees omitted were identified as "high omit" items and the remaining items were considered "low omit" items. A contingency table based on the judges omit/non-omit and the examinees' high omit/low omit items was also analyzed. This procedure was also repeated for item that 25% or more of the examinees omitted.

Results

The five judges suggested an uncorrected cutscore of 22.64, corrected to 16.30. The five scores ranged from 14.78 to 30.86 (6.48 to 26.58 corrected) with a standard deviation of 6.96 (8.70 corrected). The cutscores determined by each judge are displayed in Table 1. The reliability of the average of the judges was 0.91.

Insert Table 1 about here

Twenty-one items were identified by two or more judges as possible omit items (i.e., in their judgment, minimally competent test takers would be unable to eliminate any options), eight items were identified as such by three or more judges, and two items were identified as such by four judges. No items were identified as possible omit items by all five judges.

The examination was administered to 108 examinees who obtained a mean of 21.30 (corrected mean of 17.80) with a standard deviation of 9.40 (10.43 corrected). The internal consistency reliability of the uncorrected examination was 0.91. Thirty-two examinees were identified as "minimally competent" as their scores fell within two standard errors of measurement of the corrected scores.

For the 21 items identified as possible omit items by the judges' ratings (two or more agreeing judges), the percent of minimally competent candidates omitting the items ranged from 6 to 81. For the remaining 27 items, the percents ranged from zero to 50. The Mann-Whitney U statistics was 3.72, $p < .01$, indicating that the items identified as possible omit items were omitted at a higher rate than the items not identified as such.

Nine items were omitted by 50 percent or more of the minimally competent candidates. Eight of these nine items coincided with the judges ratings. (See Table 2) More important to notice is that of the 27 items the judges rated as non-omit items, only one was omitted by 50% or more of the minimally competent candidates.

When a "high omit" item was redefined as an item that the 25% or more of the minimally competent candidates omitted, 24 items were identified. Sixteen of these items coincided with the judges' ratings. These results are shown in Table 3. Again, the interesting result is that only 8 of the 27 items identified as low omit items by the judges were in fact omitted by more than 25% of the minimally competent test takers.

Insert Tables 2 and 3 about here

Discussion

Applying a correction for guessing formula to a cutscore derived via the Nedelsky method will penalize candidates who omit items for which they have partial knowledge. The results of this study suggest that this group of examinees tended to respond with an omit pattern to the items similar to that predicted by the judges. That is, the minimally competent examinees tended to omit those items identified as "omit" items based on the judges' ratings and not to omit those items identified as "non-omit" items from the judges ratings.

In order for the Nedelsky/correction for guessing combination to work, the judges must be able to correctly identify the items that will not be omitted by the minimally competent candidates. Of the 27 items designated as "non-omit" items from the judges' ratings, only one was omitted by more than 50% of the minimally

competent examinees and 8 were omitted by 25% or more. The data for this study suggests that, for this group of minimally competent candidates, the judges were able for the most part to predict the items that would and would not be omitted.

Of course, there are individual differences in guessing behavior, and some minimally competent examinees will omit items for which they have partial knowledge. For the minimally competent examinee who fits this mold, the cutoff score adjusted for guessing is too strict. Research is needed to examine the individual test taking behaviors of specific candidate groups in order to further clarify the appropriateness of a correction for guessing formula/Nedelsky combination. It may be necessary to direct some examinees to an alternative-elimination strategy in order to enhance their recognition of items for which they do indeed have partial information. This might be particularly useful for young or naive test takers.

The results of this study suggest that the effect of applying the correction for guessing to a Nedelsky cutscore may not be overly harsh since most of the minimally competent candidates responded as expected. Instructions which require the examinees to physically eliminate alternatives that they know are incorrect for each item may decrease the tendency for some examinees to omit items for which they have partial knowledge. This change in the test taking behavior of some examinees may increase the range of minimally

Method

11

competent candidates for whom the corrected for guessing adjusted
cutscore is appropriate.

Acknowledgements

The authors extend yet another offering of "everlasting gratitude" to the panel of judges who rated the 48 items of the Mathematics Achievement Test: Charles J. Ansorge, David Moshman, Steven L. Wise, and Joseph Witt. Anne Harvey is also thanked for her help in the data preparation.

References

- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), Educational measurement. Washington, DC: American Council on Education, pp. 514-515.
- Baggaley, A. W. (1969). Mathematics for introductory statistics. A programmed review. New York: John Wiley and Sons.
- Livingston, S. A., and Zieky, M. J. (1982). Passing scores: A Manual for setting standards of performance on educational and occupational tests. Princeton, N.J.: Educational Testing Service.

Table 1
Cutscores Suggested by Each Judge

<u>Judge</u>	<u>Raw Cutscore</u>	<u>Corrected Cutscore</u>
1	14.78	6.48
2	30.86	26.58
3	25.30	19.63
4	15.98	7.98
5	26.26	20.83

Table 2

Number of Items with Fifty Percent or More
Omits by Judges Ratings of Omit/No Omit
Two or More Judges Rating Item as an Omit Item

Percent Omit	Judges Ratings	
	Omit	No Omit
Less than 50	13	26
More than 50	8	1

Table 3

Number of Items with Twenty-five Percent or More
Omits by Judges Ratings of Omit/No Omit
Two or More Judges Rating Item as an Omit Item

Percent Omit	Judges Ratings	
	Omit	No Omit
Less than 25	5	19
More than 25	16	8

Figure 1
Algebraic Proof
Definitions

Let N = total number of items

R = expected number of correct answers

W = expected number of incorrect answers

r_1 = number of items for which examinee identified all four incorrect options as incorrect (five option item).
Probability of correct answer is 1.

r_2 = number of items for which examinee identified three incorrect options as incorrect. Probability is $1/2$.

r_3 = number of items for which examinee identified two incorrect options as incorrect. Probability is $1/3$.

r_4 = number of items for which examinee identified one incorrect option as incorrect. Probability is $1/4$.

r_5 = number of items for which examinee was unable to identify any incorrect option as incorrect. Probability is $1/5$.

Expected values of scores when adjusting Nedelsky estimate for guessing.

$$\begin{aligned}
 N &= r_1 & r_2 & r_3 & r_4 & r_5 \\
 R &= r_1 + (1/2)r_2 + (1/3)r_3 + (1/4)r_4 + (1/5)r_5 \\
 W &= (1/2)r_2 + (2/3)r_3 + (3/4)r_4 + (4/5)r_5 \\
 1/4W &= (1/8)r_2 + (1/6)r_3 + (3/16)r_4 + (1/5)r_5 \\
 R - 1/4W &= r_1 + (3/8)r_2 + (1/6)r_3 + (1/16)r_4 + 0
 \end{aligned}$$

Figure 1 (continued)

Expected value if examinee omits ($r_5 - a$) items of no knowledge type, ($0 < a < N$)

$$R = r_1 + (1/2)r_2 + (1/3)r_3 + (1/4)r_4 + (1/5)a$$

$$W = (1/2)r_2 + (2/3)r_3 + (3/4)r_4 + (4/5)a$$

$$1/4W = (1/8)r_2 + (1/6)r_3 + (3/16)r_4 + (1/5)a$$

$$R - 1/4W = r_1 + (3/8)r_2 + (1/6)r_3 + (1/16)r_4 + 0$$

Note that the expected value under this condition is equal to the expected value of the judges adjusted for guessing.

Expected value if examinee omits ($r_5 - a$), ($r_4 - b$), and ($r_3 - c$) items; that is, examinee omits items for which partial knowledge exists.

$$R = r_1 + (1/2)r_2 + (1/3)c + (1/4)b + (1/5)a$$

$$W = (1/2)r_2 + (2/3)c + (3/4)b + (4/5)a$$

$$1/4W = (1/8)r_2 + (1/6)c + (3/16)b + (1/5)a$$

$$R - 1/4W = r_1 + (3/8)r_2 + (1/6)c + (1/16)b + 0$$

Note that the expected value is smaller than the expected value of the judges adjusted for guessing by the amount: $1/6(r_3 - c) + 1/16(r_4 - b)$